Congratulations! You passed!

Grade received 80% | Latest Submission Grade 80% | To pass 80% or higher

Go to next item

1. If searching among a large number of hyperparameters, you should try values in a grid rather than random values, so that you can carry out the search more systematically and not rely on chance. True or False? 1 / 1 point

- (•) False
- ( ) True

Expand

Correct

2. In a project with limited computational resources, which three of the following hyperparameters would you choose to tune? Check all that apply. 1 / 1 point

- [x] α

  Correct. Correct. This might be the hyperparameter that most impacts the results of a model.
- [ ] β_1, β_2 in Adam.
- [ ] ϵ in Adam.
- [x] mini-batch size

  Correct. Correct. This can have a great impact on the results of the cost function, thus it is worth tuning it.
- [x] The β parameter of the momentum in gradient descent.

  Correct. Correct. This hyperparameter can increase the speed of convergence of the training, thus is worth tuning.

Expand

Correct
Great, you got all the right answers.

3. During hyperparameter search, whether you try to babysit one model ("Panda" strategy) or train a lot of models in parallel ("Caviar") is largely determined by: 1 / 1 point

- ( ) The number of hyperparameters you have to tune
- ( ) The presence of local minima (and saddle points) in your neural network
- ( ) Whether you use batch or mini-batch optimization
- (•) The amount of computational power you can access

Expand

Correct

4. Knowing that the hyperparameter α should be in the range of 0.001 and 1.0. Which of the following is the recommended way to sample a value for α? 1 / 1 point

- ( ) r = -5*np.random.rand()
  alpha = 10**r
- ( ) r = np.random.rand()
  alpha = 0.001 + r*0.999

r = 4*np.random.rand()
alpha = 10**r

r = -3*np.random.rand()
alpha = 10**r

Expand

Correct
Yes. This gives a random number between $0.001 = 10^{-3}$ and 10^{0}.

5. Once good values of hyperparameters have been found, those values should be changed if new data is added or a change in computational power occurs. True/False? 1 / 1 point

False

True

Expand

Correct
Correct. The choice of some hyperparameters such as the batch size depends on conditions such as hardware and quantity of data.

6. When using batch normalization it is OK to drop the parameter $b^{[l]}$ from the forward propagation since it will be subtracted out when we compute $\tilde{z}^{[l]} = \gamma z^{[l]}_{\text{normalize}} + \beta^{[l]}$. True/False? 1 / 1 point

False

True

Expand

Correct
Correct. Since in the normalization process the values of $z^{[l]}$ are re-centered at the origin, it is irrelevant to add the $b^{[l]}$ parameter.

7. Which of the following are true about batch normalization? 0 / 1 point

The parameters β and γ of batch normalization can't be trained using Adam or RMS prop.

One intuition behind why batch normalization works is that it helps reduce the internal covariance.

The parameter ϵ in the batch normalization formula is used to accelerate the convergence of the model.

There is a global value of γ and β that is used for all the hidden layers where batch normalization is used.

Expand

Incorrect
You did not choose an option.

8. Which of the following are true about batch normalization? 0 / 1 point

$\beta^{[l]}$ and $\gamma^{[l]}$ are hyperparameters that must be tuned by random sampling in a logarithmic scale.

When using batch normalization we introduce two new parameters $\gamma^{[l]}$, $\beta^{[l]}$ that must be "learned" or trained.

Correct
Correct. Batch normalization uses two parameters

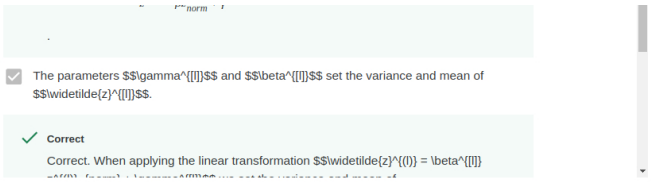
β

and

γ

to compute

$\tilde{z}^{(i)} = \beta z^{(i)} + \gamma$



The parameters $$\gamma^{[l]}$$ and $$\beta^{[l]}$$ set the variance and mean of $$\widetilde{z}^{[l]}$$.

Correct

Correct. When applying the linear transformation $$\widetilde{z}^{(l)} = \beta^{[l]}

Expand

Incorrect
You chose the extra incorrect answers.

9. A neural network is trained with Batch Norm. At test time, to evaluate the neural network on a new example you should perform the normalization using μ and σ^2 estimated using an exponentially weighted average across mini-batches seen during training. True/false? 1 / 1 point

- False
- True

Expand

Correct
Correct. This is a good practice to estimate the μ and σ^2 to use since at test time we might not be predicting over a batch of the same size, or it might even be a single example, thus using the μ and σ^2 of a single sample doesn't make sense.

10. If a project is open-source, it is a guarantee that it will remain open source in the long run and will never be modified to benefit only one company. True/False? 1 / 1 point

- False
- True

Expand

Correct
Correct. To ensure that a project will remain open source in the long run it must have a good governance body too.